Supplemental Figures

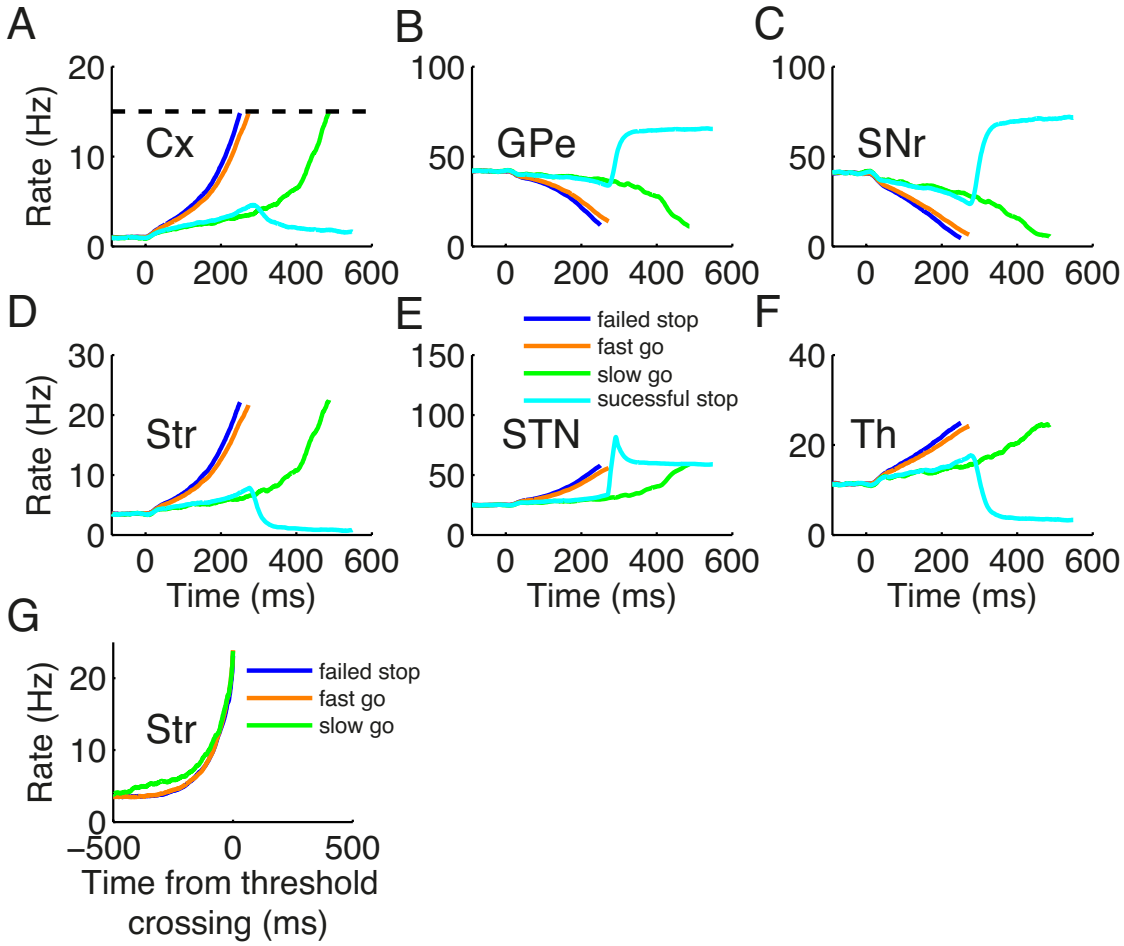

Figure S1 (related to Figure 3): Mean neuronal activities for different types of trials.
(A-F) Mean neuronal activities in each area within the circuit, aligned with the onset of go stimulus. (G) Mean neuronal activities in the Str, aligned with the time for threshold crossing in the Cx. The blue, orange, green and cyan curves represent failed stop signal trials, latency-matched fast go trials, latency-matched slow go trials, and successful stop signal trials. The latency-matched fast/slow go trials are go trials with RTs shorter/longer than SSD+SSRT, where SSD = 270 ms and SSRT =155 ms, as estimated in Fig. 3.

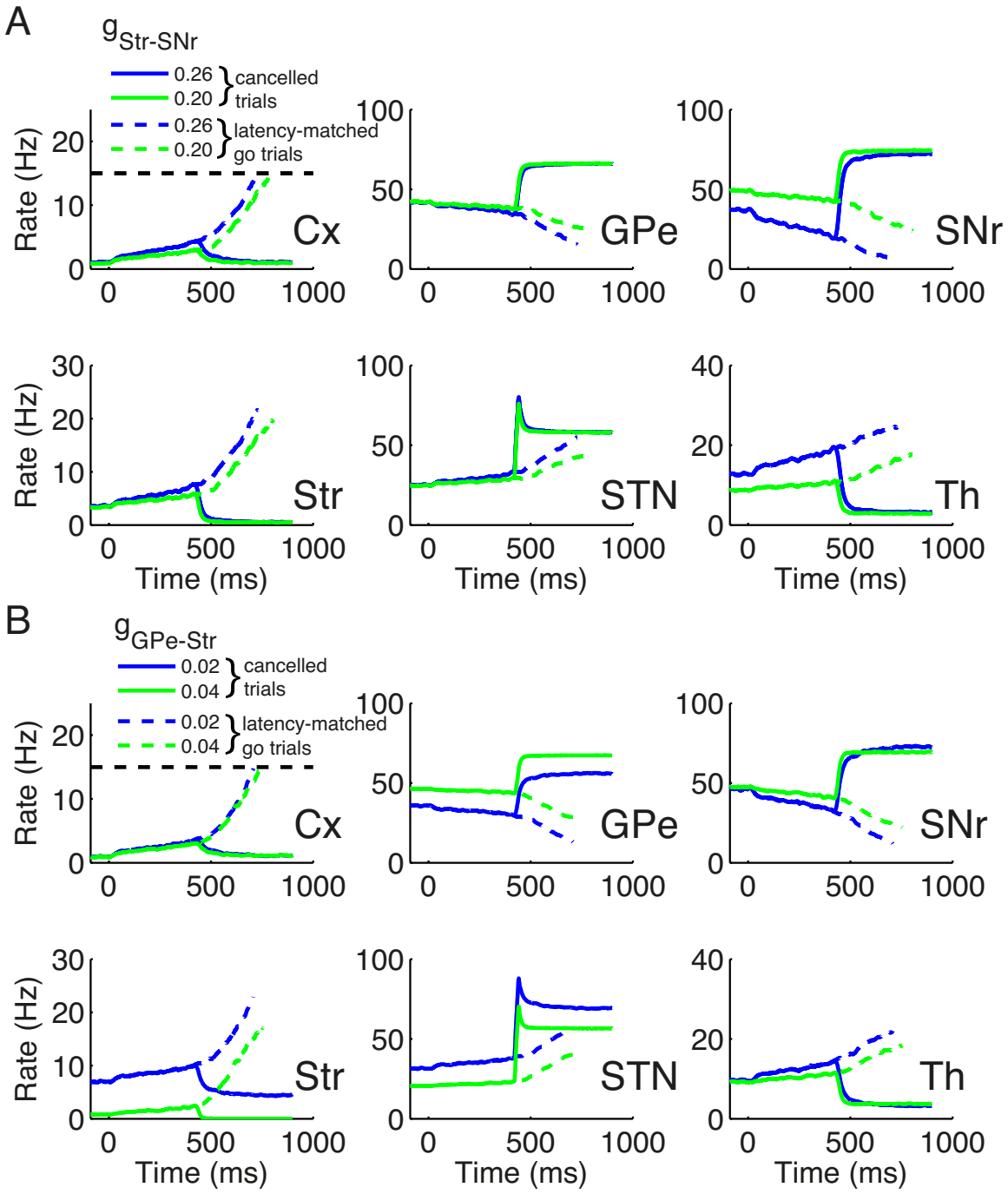

Figure S2 (related to Figure 4): **Dependence of mean neuronal activities for successful stop signal trials and latency-matched go trials on $g_{Str-SNr}$ and $g_{GPe-Str}$.** (A) Mean neuronal activities in each area within the circuit with $g_{Str-SNr}$= 0.20 nS (green) and 0.26 nS (blue). (B) Mean neuronal activities in each area within the circuit with $g_{GPe-Str}$= 0.02 nS (blue) and 0.04 nS (green). The solid and dashed curves represent successful stop signal trials and latency-matched go trials, respectively. The mean activities are obtained from averaging over 200 trials of the same type. Note that in (B), for fair comparison the spontaneous firing rate of the SNr is fixed by adjusting the external input rate υ_{ext} to it.

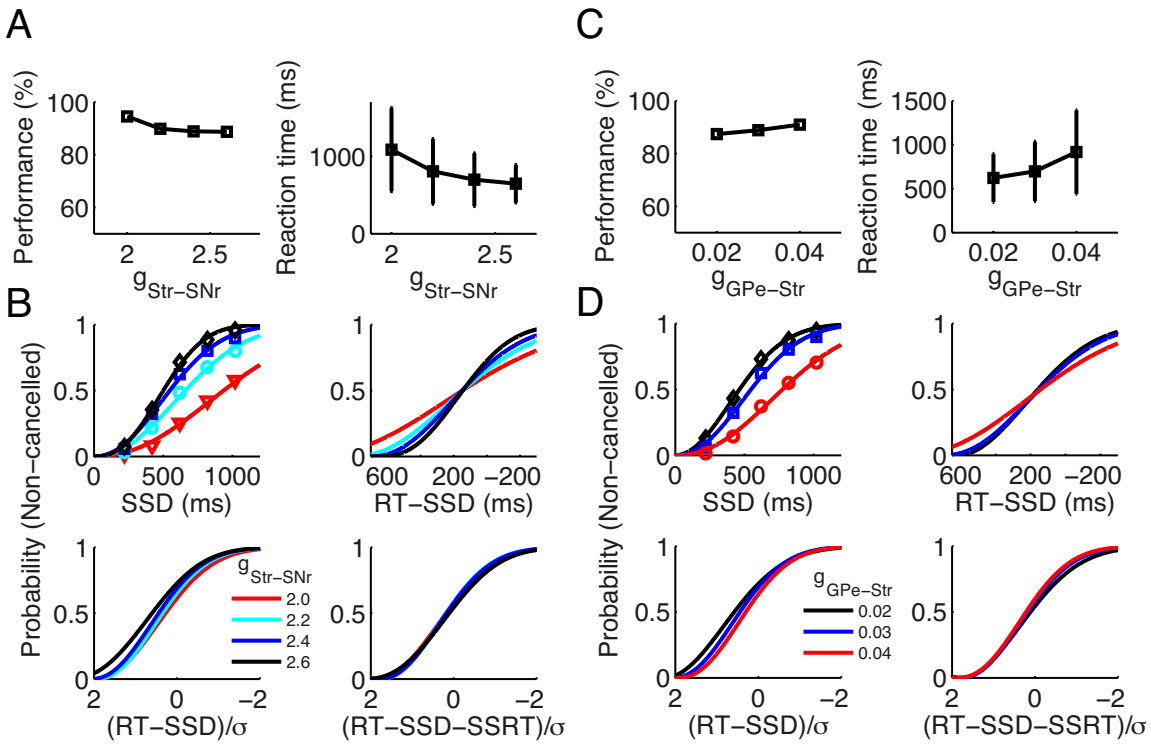

Figure S3 (related to Figure 4): Performances and RTs of the go trials and alignment of inhibition functions when $g_{Str-SNr}$ and $g_{GPe-Str}$ are modulated. (A) Performance (Left), and mean and s.d. of RTs of go trials for different $g_{Str-SNr}$. (B) Non-cancelled probabilities for different $g_{Str-SNr}$ as a function of SSD, RT-SSD, (RT-SSD)/σ, and (RT-SSD-SSRT)/σ, respectively. (C-D) Same as (A-B), but for the modulation of $g_{GPe-Str}$. Here RT and σ are mean and s.d. of the RTs of go trials. The inhibition functions for different $g_{Str-SNr}$ ($g_{GPe-Str}$) are aligned when the dependence on SSD is transformed to the dependence on the z score of the relative finishing time (ZRFT), (RT-SSD-SSRT)/σ.

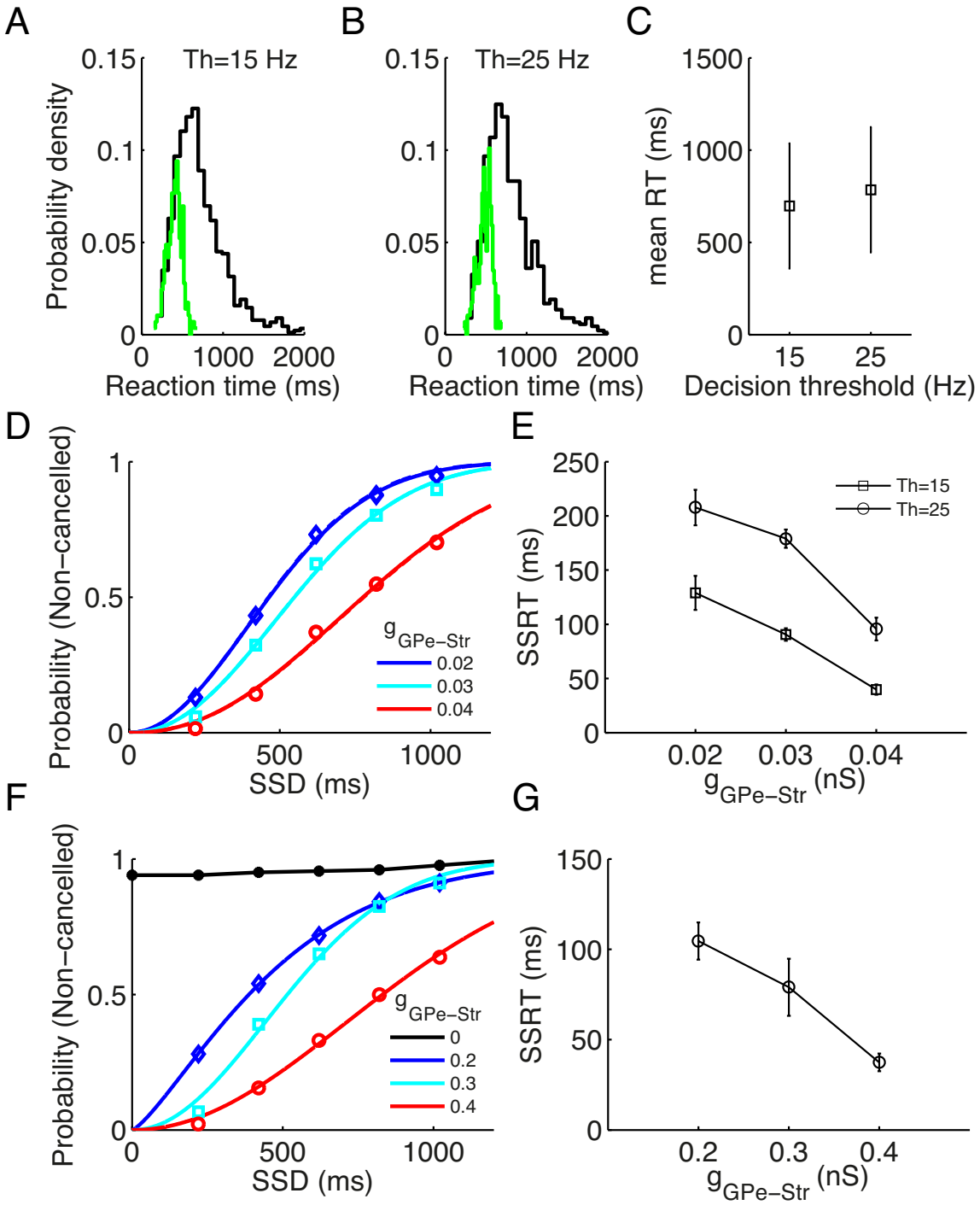

Figure S4 (related to Figure 4): Impact of back-projection from the GPe to the Str for different levels of decision threshold and in a scaled circuit with all-to-all connections in all the areas. (A-B) RTs for go trials and stop signal trials with SSD=420 ms for decision threshold (Th) at 15 Hz (A) and 25 Hz (B). (C) Mean and s.d. for the go trials RTs. (D) Inhibition functions for different GPe-to-Str projection strength ($g_{GPe-Str}$) with Th = 15 Hz (solid lines) and Th = 25 Hz (dashed lines). Note that the solid and dashed lines are almost indistinguishable. (E) Modulation of SSRT by $g_{GPe-Str}$ shows similar trends for different levels of decision threshold. (F-G) Dependence of inhibition functions and SSRT on $g_{GPe-Str}$ in a scaled circuit with all-to-all connections in all the areas. In (F-G), the numbers of neurons in the GPe and STN are scaled

down from 5000 in original model to 500. The connection probabilities within the STN-GPe sub-circuit are all scaled to 1 now, such that the connection strengths are scaled by a factor 10*p, where p is the connection probability for the corresponding synaptic connection in the original model. Error bars indicate s.d. of SSRT values estimated for each SSD. The inhibition functions in (D) and (F) are fitted by the Weibull function.

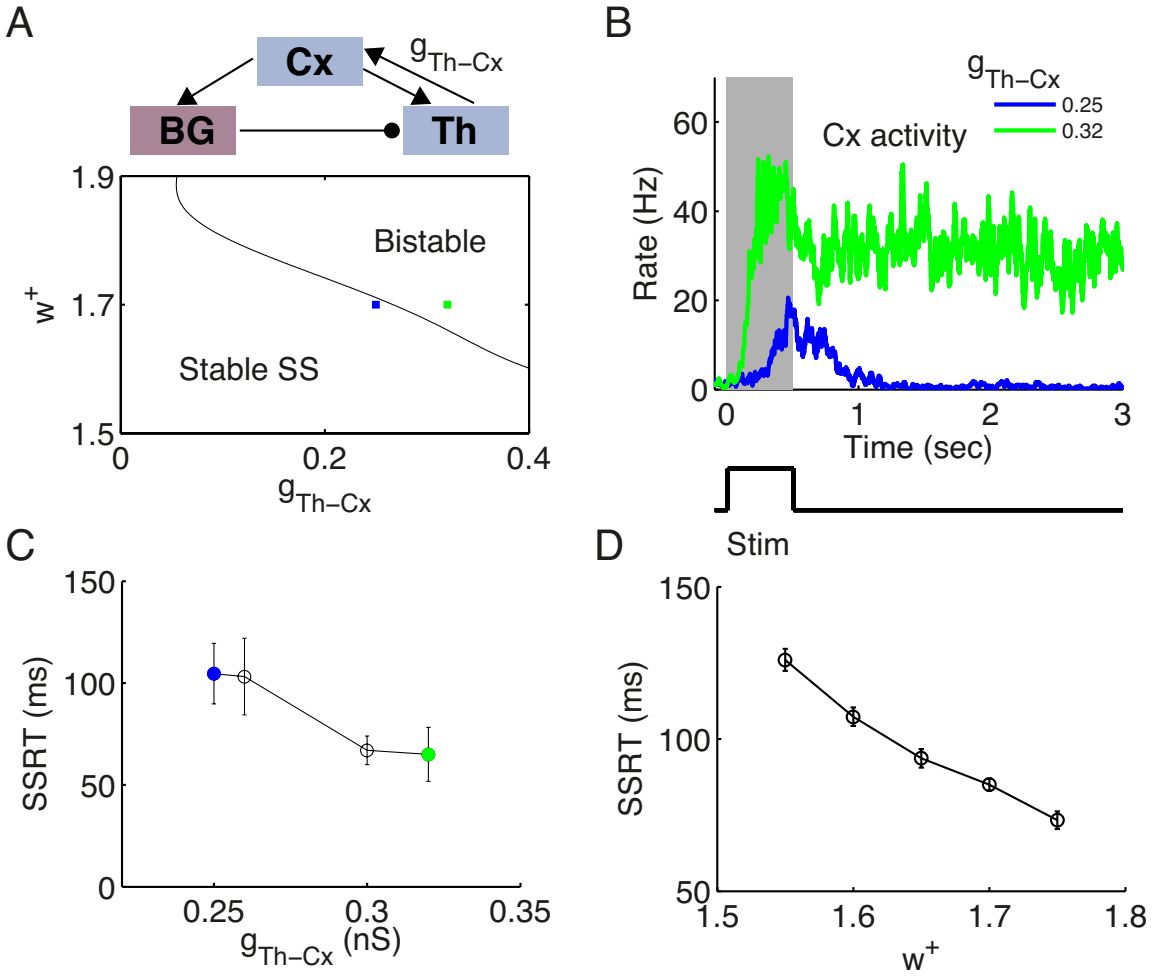

Figure S5 (related to Figure 5): Dependence of SSRT on thalamocortical connection strength when there is a corticothalamic sub-loop and on local cortical recurrent connection strength w^+. (A) Upper panel: schematic of the Cx-BG-Th loop with a Cx-Th sub-loop. Lower panel: the state diagram as a function of the local recurrent feedback strength w^+ in the Cx and the subcortico-cortical loop feedback strength g_{Th-Cx}. Stable SS: region where only the spontaneous state (SS) is stable. Bistable: region where the SS and target-selective elevated persistent activity state are stable. Two sample networks are indicated in the parameter space at g_{Th-Cx}=0.25 nS (blue) and 0.32 nS (green) with w^+=1.7. (B) Neuronal activity of the winner population in the Cx in response to a brief stimulus. The green and blue curves correspond to the green and blue squares in (A). The gray-shaded area shows the neuronal activities during stimulus onset, which lasts for 0.5 sec (lower panel). (C) SSRT shows a two-level modulation by g_{Th-Cx}, corresponding to whether there exists persistent activity or not. (D) SSRT decrease monotonically when w^+ is increased. In (A-C), w^+ = 1.7, g_{Cx-Th} = 0.2 nS. In (D), g_{Th-Cx} = 0.32 nS, g_{Cx-Th} = 0. Error bars indicate s.d. of SSRT values estimated for different SSD.

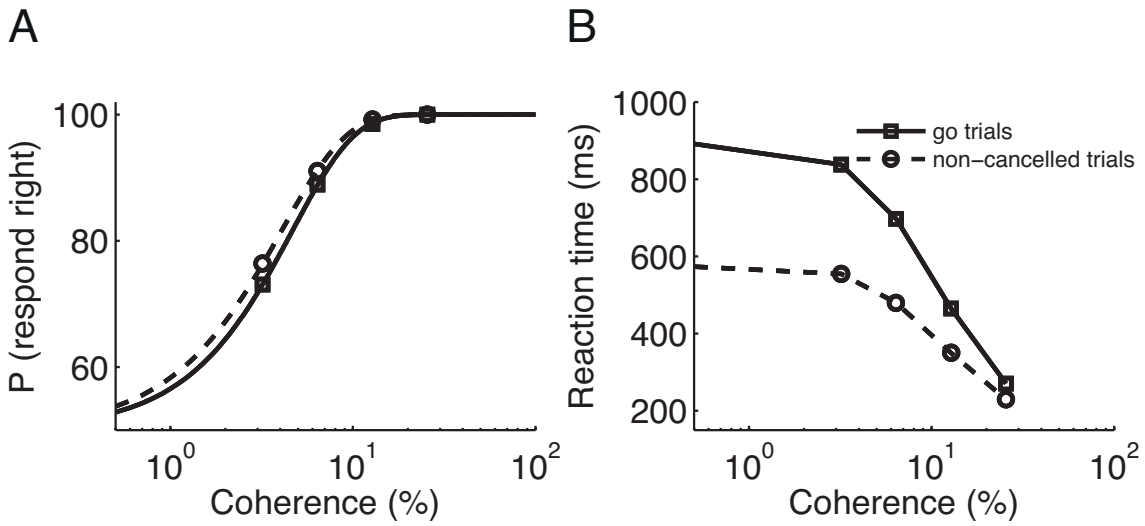

Figure S6 (related to Figure 6): Probabilities of choosing right (A) and mean RTs (B) in go trials and non-cancelled stop signal trials when the net direction of random-dots stimulus is right with different coherences. For non-cancelled stop signal trials, the probability of choosing right increases and mean reaction time decreases with the increase of motion coherence, which are similar to that for go trials. Note that in experiments the SSD is usually adjusted following a staircase procedure to ensure total non-cancelled probability around 0.5. In modeling study we performed simulations for each SSD, and here the performances and mean RTs for non-cancelled trials are average values over two SSDs that have non-cancelled probabilities close to 0.5 for each coherence level of random-dots.

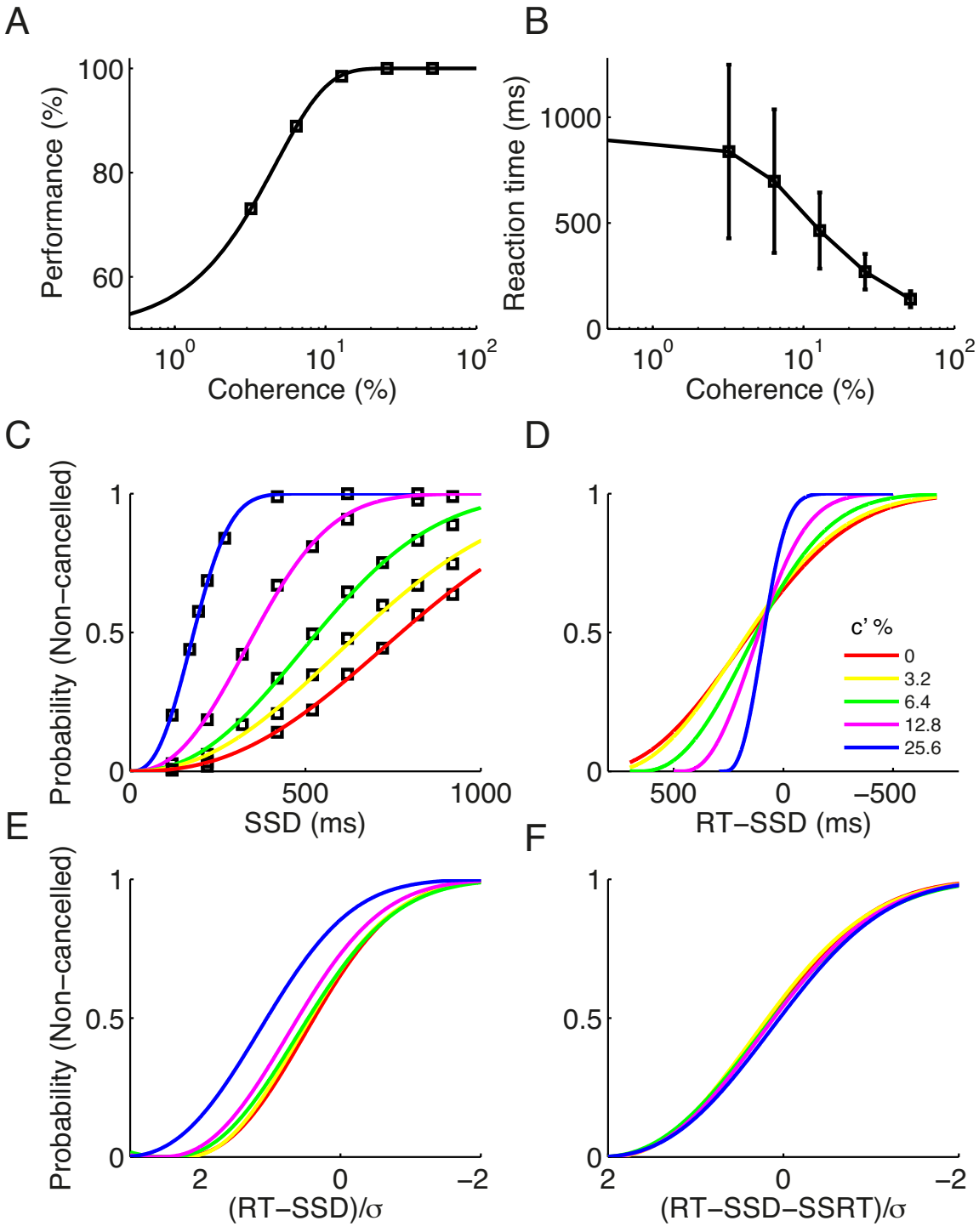

Figure S7 (related to Figure 6): Performances and RTs of the go trials and alignment of inhibition functions when the motion coherence c' is varied. Similar to Figure S3, but for the variation of c'. The inhibition functions for different c' are aligned when the dependence on SSD is transformed to the dependence on ZRFT, (RT-SSD-SSRT)/σ.

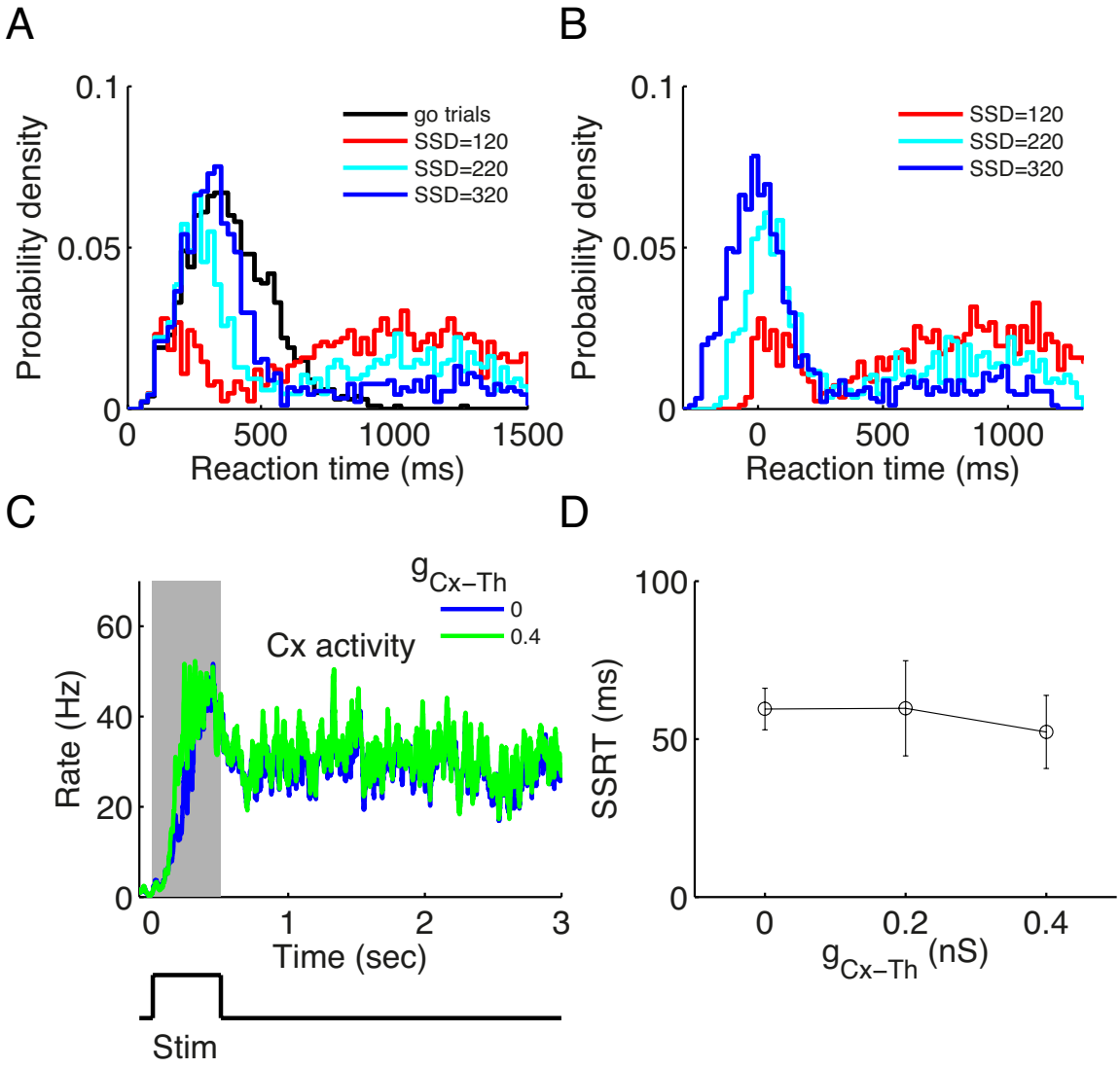

Figure S8 (related to Figure 7): Impact of corticothalamic sub-loop on RT distributions and SSRT. (A-B) RT distributions for the go trials and non-cancelled stop signal trials aligned with the go stimulus onset (A) and stop signal onset (B). (C) Neuronal activities of the winner population in the Cx for $g_{Cx-Th} = 0$ (blue) and $g_{Cx-Th} = 0.4$ nS (green), respectively. The Cx circuit shows persistent activity in both cases. The gray-shaded area shows the neuronal activities during stimulus onset, which lasts for 0.5 sec (lower panel). The stimulus is a random-dot with c'= 25.6%. (D) SSRT is insensitive to g_{Cx-Th}. Error bars indicate s.d. of SSRT values estimated for different SSD. In (A-B), $g_{Cx-Th} = 0.4$ nS, g_{Th-Cx} = 0.26 nS. In (C-D), g_{Th-Cx} = 0.32 nS.

Supplemental Experimental Procedures

Behavioral task simulation

We simulated a two-alternative choice task and considered two types of stimulus. The go stimulus is implemented by an increase of the rate of external background inputs, modeled as Poissonian spike trains, to the selective population (left or right) in the cortical circuit from 2.4 kHz to 2.464 kHz. For the random-dots stimulus, the direction of coherent motion is assumed to be left- or rightward. Two signals, representing sensory input induced by the leftward and rightward moving dots, are fed into left- and right-preferring cortical modules Cx_L and Cx_R, respectively. The firing times of each input neuron are sampled from a Poisson process with a mean rate μ. The mean μ depends linearly on the dot coherence level and is given by the equations $\mu_0 + \mu_A \times c'$ for the preferred direction and $\mu_0 - \mu_B \times c'$ for the non-preferred direction, where μ_0 (20 Hz) is the baseline input, c' (between 0% and 100%) is the coherence level, and μ_A (60 Hz) and μ_B (20 Hz) are proportionality factors (Wang, 2002). The stop signal is simplified as a constant input to the STN, entering the BG via the hyperdirect pathway. The stop signal is implemented by an increase of the external Poisson input rate to all the STN neurons from 4.0 kHz to 5.2 kHz.

A correct go trial or failed (non-cancelled) stop signal trial is defined as a trial in which the firing rate of the selective population in the Cx crosses a threshold at 15 Hz. A successful stop signal trial is defined as a trial in which the firing rate of the selective population in the Cx never crosses the threshold. Note that the results in the paper do not depend on the choice of threshold value in the Cx. From Figure S4 (A-C), we saw that the mean RT distribution is shifted toward right for a higher threshold, but the standard deviation is not changed. The inhibition function is not influenced by the choice of the decision threshold (Figure S4, D). This is expected, since whether a trial is cancelled or not by a stop signal is determined much earlier than the time for threshold crossing where the stop signal is unlikely to cancel a trial due to the attractor dynamics. We saw that the SSRT is increased for a higher threshold, since the mean RT for go trials is increased and the inhibition function is not changed (the SSRT is roughly given by the mean RT for go trials subtracted by the SSD corresponding to 50% non-cancelled probability). The trend of SSRT modulation by model parameters, illustrated here by the strength of GPe-to-striatum back-projection, is conserved (Figure S4, E). When there is no stop signal, model performance is defined as the fraction of correct go trials. In stop signal trials, the non-cancelled probability is defined as the ratio between the fraction of non-cancelled stop signal trials and the performance when there is no stop signal. This definition ensures that the non-cancelled probability approaches to 1 for large SSD.

Integration method for estimating the SSRT

The integration method has been widely used in estimating the SSRT, which is illustrated in Figure 3A. When the cumulative density function (cdf) of RTs is plotted for go trials (black) and non-cancelled stop signal trials (green) in Figure 3A, the crossing point of the horizontal dashed line (representing the total non-cancelled probability) with the black curve gives an estimate of the SSD plus SSRT (Figure 3A, upper panel). Below this time (red vertical dashed line), the probabilities for the RTs of go trials and non-cancelled trials roughly coincide with each other (Figure 3A, lower panel).

In Figure 7A, a modified integration method as introduced in Mayse et al. (2014) was illustrated. The intersection of the cdf for the RTs of the non-cancelled stop signal trials (scaled with the non-cancelled probability; green solid curve) with the 99.9% confidence interval of the cdf for the RTs of go trials (black dashed curve) gives an estimate of the SSD plus the SSRT.

Neuron model

We used the leaky integrate-and-fire model for all the neurons. The membrane dynamics is given by

$$C\frac{dV}{dt} = -g_L(V - V_L) - I_{syn}\,, \tag{1}$$

where C = 0.5 nF, g_L= 25 nS, V_L = -70 mV. When the membrane potential reaches a boundary V_b, it is reset to V_r, where V_b = -50 mV and V_r = -55 mV. The synaptic current I_{syn} is given by

$$I_{syn} = g_1 s(V - V_E) + \frac{g_2 s(V - V_E)}{1 + e^{-0.062V/3.57}} + g_3 s(V - V_I)\,, \tag{2}$$

where the reversal potentials V_E = 0 and V_I = -70 mV. g_k is the synaptic efficacy, where the indices k=1, 2, 3 indicate AMPA, NMDA and $GABA_A$ synapses, respectively. The gating variable s satisfies

$$\frac{ds}{dt} = \sum_j \delta(t - t^j) - \frac{s}{\tau} \quad (3)$$

for AMPA and $GABA_A$ receptor mediated currents and

$$\frac{ds}{dt} = \alpha(1 - s)\sum_j \delta(t - t^j) - \frac{s}{\tau} \quad (4)$$

for NMDA receptor mediated current, where t^j is the time for the jth spike and $\alpha = 0.63$. For the decay time constant τ, we use 2 ms for AMPA, 5 ms for $GABA_A$, and 100 ms for NMDA mediated currents. I_{syn} describes both the synaptic inputs from other neurons within the circuit and background inputs modeled as Poissonian spike trains representing sources beyond the circuit. We use a short time delay, 0.2 ms, for synaptic transmission.

To soften the positive feedback of the cortico-subcortico-cortical loop, we introduced short-term depression (STD) at thalamocortical synapses, but the results do not depend on this assumption. The STD is implemented by including a factor D that multiplies all terms on the right side of Eq. (2) (Hempel et al., 2000), which satisfies

$$\frac{dD}{dt} = -pD\sum_j \delta(t - t^j) + \frac{1-D}{\tau_D} \quad (5)$$

where p = 0.45 and τ_D = 600 ms.

Network structure

Each BG nucleus includes two populations of neurons selective to left and right targets or moving dots, respectively. Following Wei et al. (2015), each population in the Str and SNr includes 250 neurons with all-to-all connections, while each population in the STN and GPe includes 2500 neurons with sparse connection probabilities: 0.05 for STN→GPe, 0.05 for GPe→GPe, and 0.02 for GPe→STN. There are no known recurrent connections within the STN (Sato et al., 2000), hence such connections are not included. All other projections are all-to-all for the left and right selective populations, respectively. The sparse connections in the STN-GPe sub-circuit are inherited from our previous model (Wei et al., 2015), where we proposed a mechanism for excessive beta oscillation in Parkinson's disease that depends on the sparse connections of the STN-GPe circuit. Here we adapted this structure with the hope to provide a framework allowing for investigating such a parkinsonian state in the future. To confirm that the sparse connectivity in the STN-GPe circuit is not essential, we performed additional simulations (Figure S4, F-G), where we scaled the STN and GPe circuits to have all-to-all connections and 500 neurons in both areas. We observed qualitatively the same results as in the original model, illustrated by the impact of GPe-to-Str back-projection strength. Similarly, the thalamus includes also two populations due to the topographic connections from the SNr to higher order thalamus (Gulcebi et al., 2012), with each population including 250 neurons with all-to-all connections.

The connection efficacies used in the model are the following (in unit of nS): g_{Cx-Str} = 1.0, $g_{Str-Str}$ = 1.0, $g_{Str-SNr}$ = 2.4, $g_{Str-GPe}$ = 3.0, $g_{GPe-GPe}$ = 1.5, $g_{GPe-Str}$ = 0.03, $g_{GPe-STN}$ = 0.8, $g_{GPe-SNr}$ = 0.04, $g_{STN-GPe}^{NMDA}$ = 2.0, $g_{STN-SNr}$ = 0.06, g_{SNr-Th} = 0.09. We will use g_{Th-Cx} = 0.32 nS, unless otherwise indicated. The corticothalamic strength g_{Cx-Th} is taken to be zero except in the last section of Results. For the cortical network, we used the same synaptic connection properties as in past studies (Lo and Wang, 2006; Wang, 2002), except that the efficacy of background inputs to the two selective populations are reduced from 2.1 nS to 2.0 nS. This modification weakens the cortical background input to support a distributed attractor state.

Background inputs, which are incorporated into equation (4), are described as Poissonian spike trains. All of the BG nuclei in the model receive AMPAergic background inputs, with possible cortical and thalamic origins. GPe neurons also receive GABAergic background inputs representing additional inputs from the striatum. The synaptic efficacies and Poisson rates used for the background inputs in our model are (in unit of nS and kHz, respectively): 4.0 and 1.6 for the Str; 2.0 and 4.0 for GPe AMPA receptors; 2.0 and 2.0 for GPe $GABA_A$ receptors; 1.6 and 4.0 for the STN; 6.0 and 0.8 for the SNr, 2.0 and 3.2 for the Th. Note that in Figure 4 (B, D), for fair comparison when $g_{GPe-Str}$ or $g_{Ark-Str}$ is varied, the spontaneous firing rate of the SNr is fixed by adjusting the external Poisson input rate υ_{ext} to it. For $g_{GPe-Str}$ ($g_{Ark-Str}$) taking the values 0, 0.02, 0.03 and 0.04 nS, the values of υ_{ext} used are 1.1, 0.94, 0.8 and 0.76 kHz, respectively.